

How We See

Richard A. Andersen
California Institute of Technology

ABSTRACT

The visual world is imaged on the retinas of our eyes. However, “seeing” is not a result of neural functions within the eyes but rather a result of what the brain does with those images. Our visual perceptions are produced by parts of the cerebral cortex dedicated to vision. Although our visual awareness appears unitary, different parts of the cortex analyze color, shape, motion, and depth information. There are also special mechanisms for visual attention, spatial awareness, and the control of actions under visual guidance. Often lesions from stroke or other neurological diseases will impair one of these subsystems, leading to unusual deficits such as the inability to recognize faces, the loss of awareness of half of visual space, or the inability to see motion or color.

INTRODUCTION

When we look around us, seeing is so effortless that we think we naturally perceive what is actually out there in the world. But, in fact, the brain works very hard at reconstructing its own reality – what we refer to as neural representation. In the well-known Kanizsa triangle (Figure 1), you can see illusory contours that are created by the occlusions, the lines, and the little Pac-Man figures. These “contours,” and the perceived variations in brightness lie entirely within your brain and do not exist in the real physical world. Because the brain is often faced with an ambiguous, ill-defined environment, it is very useful to reconstruct such lines. Our brains have to make hypotheses about what we think is in the real world. We hope that these hypotheses are right, because a mistake could be fatal. In terms of evolutionary pressure, the brain has evolved over time to create its own reality that meshes with the world in such a way as to enable the organism to survive.

Neurobiologists believe that at least one-third of our approximately one hundred thousand genes are exclusively involved in brain function. With that limited number of genes, we cannot completely specify all the complex



Driving down a street generates an optic flow of motion signals, where surrounding objects seem to radiate outward from the focus point. These complex signals are processed in the higher levels of the brain.

Photo by David Bradley

connections and structure in the brain, so during some periods of development, the brain has to look to the outside world for assistance in forming its structures. At a very early age, for example, the brain becomes plastic for vision; during this critical period information from the two eyes, which compete with each other, is used to actually set up the appropriate neural machinery for depth perception.

The understanding that the visual system actually constructs images of reality has led to an exciting revolution not only in neuroscience but also in the field of philosophy. A new school of philosophy called neurophilosophy has reconsidered what the nature of reality and the nature of knowledge are, based on what we now know about how the brain works and about the changes that occur in our neural networks over the course of development.

Visual information required to construct this representation of “Where” system the world comes in through the eyes and is projected on the retina; then the optic nerve sends this information to the thalamus, which passes it up to the primary visual cortex (called V1), where simple aspects of the visual scene are first analyzed. Then information is projected out to cortical areas around the primary visual cortex, and they process the visual image

Author's Current Address:
California Institute of Technology, Biology Division 216-76, 1200 E. California Boulevard,
Pasadena, CA 91125, USA.

This contribution, originally presented at our 1998 Big Sky Aerospace Conference, is to inform those readers to whom the subject may be unfamiliar and to provide a benchmark for readers working in this field.

0885/8985/08/ USA \$25.00 © 2008 IEEE

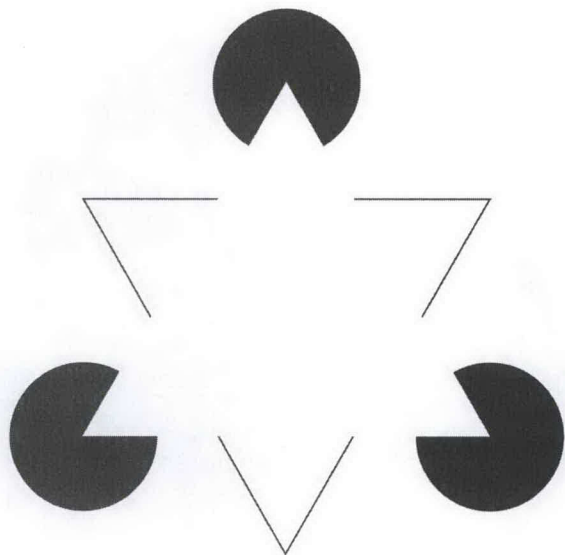


Fig. 1. Kanizsa Triangle

more elaborately; here is where the more complicated cognitive functions take place. The information travels along two processing streams – one to the upper part of the brain and the other to the lower part. In 1982, two neuropsychologists from the NIMH, Mort Mishkin and Leslie Ungerleider, proposed that the pathway to the upper part of the brain was the “where” pathway, which tells us the location of an object (Figure 2). They labeled the lower route the “what” pathway, because it seems to handle information about the object itself. Patients with injuries, or lesions, to the upper pathway can identify objects and the differences between objects, but cannot tell where they are. With lesions to the lower area, a person can tell where things are but cannot identify them. Lesions in this area can cause an interesting syndrome called prosopagnosia, in which people cannot identify faces, including their own. This object-based pathway is also important for the perception of color.

A typical lesion in the upper, or “where,” pathway might leave a patient unable to pour a liquid into a glass. He can see the glass and he knows it is a glass, but he cannot figure out where the glass is with respect to his body. Another one of the deficits from damage to this pathway is the inability to attend to the area of space opposite to the hemisphere that was damaged.

Monkeys have visual functions similar to ours. They see color the way we do; they see motion and depth; they perceive objects; they make eye movements in the same ways that we do. So they make ideal animal models for studying the human brain, because we can do experiments with monkeys that we obviously cannot do with people. We have several rhesus monkeys who participate in experiments for a period of years. Recently we have been successful in placing them in zoos for their retirement. A common technique for studying the visual system introduces very fine (about the diameter of a human hair) wire electrodes into a monkey’s cerebral cortex. We park these electrodes near cortical

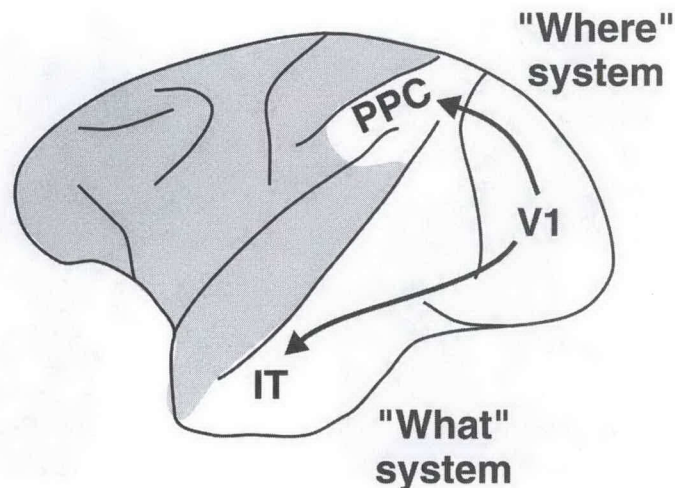


Fig. 2. Schematic of the “what” and “where” pathways

neurons. During the experiments the monkeys are awake and performing different tasks that they were trained to do, such as moving their eyes toward a stimulus, reaching toward a target, or pressing a button for a juice reward. In this way the monkeys “tell” us what they see. As they do their tasks, the electrodes record the activity of the nerve cells. Then we can correlate the activity of specific cells with the behaviors or perceptual experiences the animals have.

Figure 3 shows the type of signal that we record on one of these electrodes. Time is plotted along the x-axis, while the y-axis displays the membrane potential, or electrical activity, coming from one of these nerve cells. When we shine a light or present a stimulus to the animal, a cell that is involved in the perception of that stimulus begins to fire action potentials – pulses that are the communication method for nerve cells. These signals will then be transferred via synapses to other nerve cells to which this nerve cell projects. This synaptic transmission is how messages get sent through the cerebral cortex, and by tapping into this system with our electrodes, we can determine the locations of very specific types of visual processing that the brain uses to reconstruct reality.

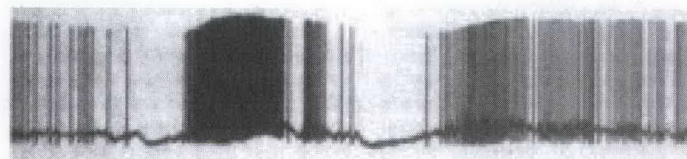


Fig. 3. A record of action potentials

The brain uses five basic strategies in its visual processing: population coding, functional localization, parallel processing, hierarchical processing, and association. A single neuron in the brain looks at only a small piece of the world. This fragmentation actually starts in the retina, which has the image of the whole visual field on it, yet a single cell receives its input from only a tiny part of that image. So we



Fig. 4. The 1961 Rose Bowl prank is an example of population coding

have to realize that each time we record a signal from one of these nerve cells, we're seeing only a small part of the entire visual message. This brings us to a concept known as population coding – the idea that a whole perception is stored across many units. Our brains are a bit like TV sets; we can think of neurons as corresponding to the pixels on the screen. Of course, a normal TV screen measures about 600×400 pixels, while the brain contains about a hundred billion cells. Each one of these cells can change its activity over a certain range to store a small bit of the “picture.” A simple example of population coding can be seen in the great Rose Bowl prank of 1961, where each University of Washington fan knew only that he or she was holding up a white or a dark card and, fortunately for the Caltech students who pulled off the prank, no one person could see the whole message (Figure 4). When they all flipped their cards in unison, they inadvertently spelled out CALTECH. To understand the brain and how it processes visual images, we not only have to know what each single element is saying, but also what the whole ensemble of activity is saying together.

A second important feature of how the brain works is known as functional localization. This concept refers to the fact that different parts of the cortex are specialists in particular visual processes. At the turn of the century, a German neuroanatomist, Korbinian Brodmann, divided the human brain into about 50 different areas simply by looking at sections of it under a microscope and noticing the differences in nerve-cell structure or packing density in different cortical regions. With the advent of microelectrode recording techniques, neurophysiologists in the 1970s began dividing the brain up into areas based on different functional activities as well. Often these functional areas corresponded to Brodmann's anatomical ones; for example, V1 was his area 17. But others, like Brodmann's area 19, turned out to contain many different cortical areas delineated by functional differences. It is also important, in dividing up the cortex, to notice that one area might connect to some areas and not to others, so that different cortical areas have specific

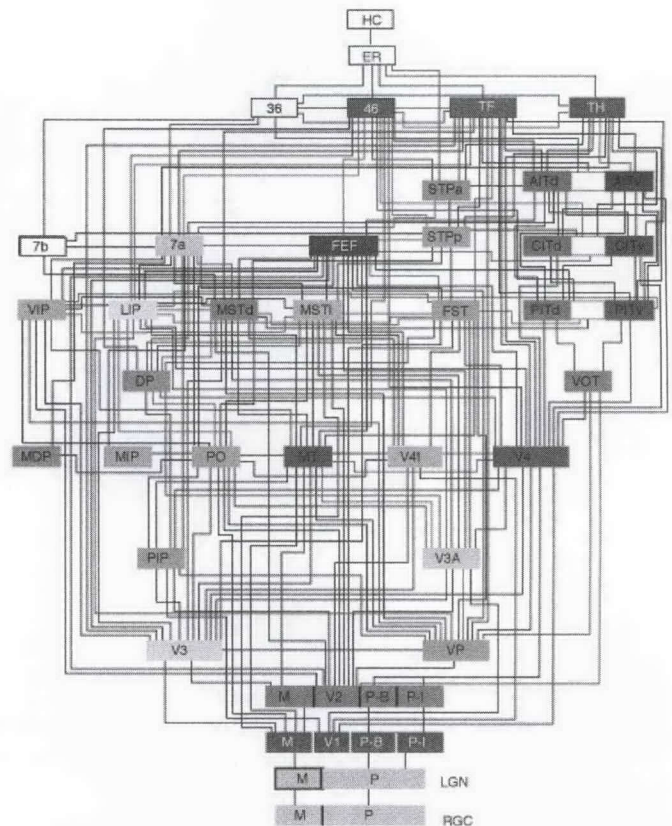


Fig. 5. Schematic of the hierarchy of cortical visual areas

connectivities between them. About 35 different cortical areas have been identified as being involved with vision in monkeys, and there are probably even more in our own brains.

Each box in Figure 5 (created by Dan Felleman and David Van Essen) corresponds to a cortical area that has a particular function. The primary visual cortex is at the bottom, and information eventually rises to the highest levels of processing in the association cortex, which then connects to the motor cortex to direct movements. The areas on the left correspond to the “where” pathway, and those on the right to the “what” pathway.

Of the three remaining strategies, parallel processing divides up information and processes it in parallel along separate lines, and hierarchical processing transfers information from one level to another through more complicated analyses as it moves up the system. The final important concept is association – after we have broken up the image and analyzed it along parallel and hierarchical lines, ultimately, we have to combine it again into a single perception.

Parallel processing streams break up and analyze different aspects of a scene. For example, when we see a red bouncing ball, we perceive it as one thing – a red bouncing ball. But in our brains some areas are processing the red, others are simultaneously processing the spherical shape, and others are processing its motion. In the last 10 years, we discovered that visual information is immediately divided in the primary

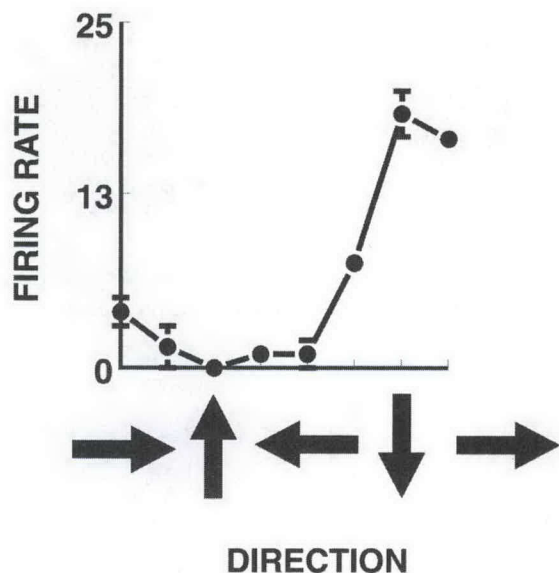


Fig. 6. Data from an MT neuron, which shows strong preference for downward motion

visual cortex (V1) into parallel streams. For example, within V1 are some repeating little patches, recently discovered and imaginatively referred to as “blobs,” which contain concentrations of nerve cells that are sensitive to color. These cells preferentially project to a particular area of V2 called the “thin stripe” area. These thin stripes are involved in color processing, and they project, in turn, into an area called V4, which is also specialized for color. If area V4 sustains damage, the patient will have difficulty perceiving color in the opposite visual field. Another pathway – for motion processing – goes from V1 to the “thick stripe” region of V2, and then onto an area called V5 or MT (for medial temporal area, discovered by John Allman, Caltech’s Hixon Professor of Psychobiology and professor of biology) in the visual association cortex. An injury to this pathway produces a very specific motion deficit; a person looking at traffic, for example, could see the cars, but would be unable to see that they are moving.

Each of these parallel streams is also organized hierarchically. Take, for example, the pathway for motion just mentioned. Cells in V1 extract some very basic information about the direction of motion, which is maintained in area MT and shown in the recording from an MT cell in Figure 6. When something moves up within the cell’s receptive field, this particular cell gives a small response, but it gives a much more vigorous response when something moves downward. This cell is giving information about the direction in which something moves, a very simple and basic sort of motion. Our recent research has tested how monkeys perceive three-dimensional structure from an object’s motion. If we were to paint little dots on a hollow, glass cylinder and view it with one eye, the cylinder would look simply like a set of dots until we turn it; then the three-dimensional shape of the glass would immediately pop out. So motion signals can give us impressions of

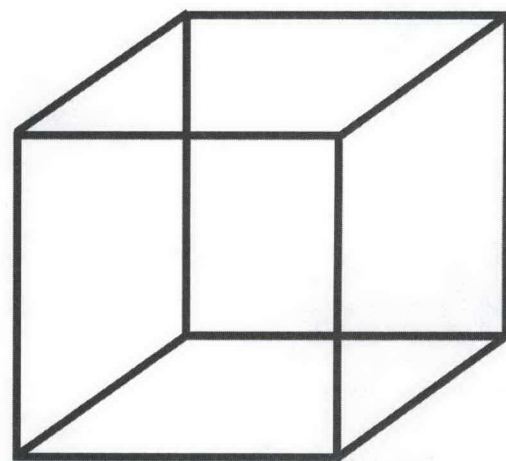


Fig. 7. Necker cube has ambiguous front and back

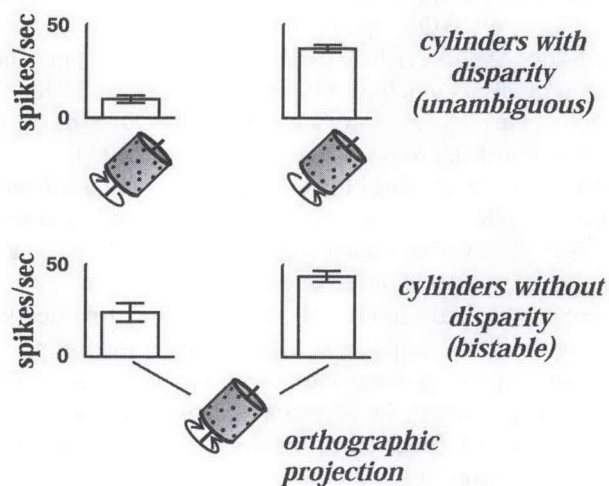


Fig. 8. Activity of MT cells varies with perceived structure of the cylinder

three-dimensional shape. Instead of using a glass with painted dots, however, we use high-speed, computer animation to generate these 3-D structure-from-motion stimuli. When we project such an image onto a flat computer monitor screen, we lose the depth information that we would normally get from looking at the cylinder with two eyes, but, amazingly, due to the motion signals, we can still perceive a revolving hollow cylinder. This computer simulation demonstrates that the brain is able to use motion signals to reconstruct three-dimensional depth. It is most interesting, however, that, since there is no depth information contained in the projected stimulus, the direction in which the cylinder appears to be rotating is ambiguous. Sometimes you may see it rotating clockwise, other times, counterclockwise. And it appears to shift directions; we refer to this spontaneous shifting as a bistable percept. An example of another bistable percept is illustrated in Figure 7: the well-known Necker cube illusion. Some people will see the upper square as being in front, and others will see the bottom square in front. If you look at it for awhile, you will see it flip spontaneously.

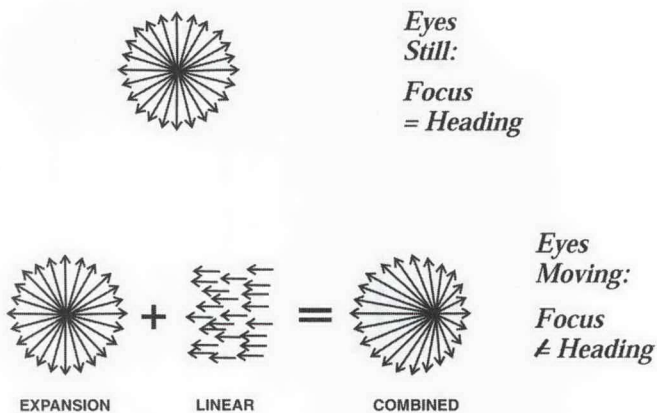


Fig. 9. Self-motion generates expanding images. Eye movements shift the expansion focus

(Sometimes it helps if you concentrate on one point and then on another to see the flipping.)

Postdoc David Bradley, grad student Grace Chang, and I trained monkeys to tell us with eye movements which direction they saw the cylinders rotating; we then recorded from their MT neurons. In some trials we added in stereoscopic depth cues in the computer display using an anaglyph technique similar to that used in the old 3-D movies of the 1950s. We found that when the monkey looks at a rotating cylinder with depth cues, the cylinders are unambiguous, and certain cells will prefer certain directions of rotation. For example, when the cylinder is rotating clockwise, it will generate a lot of activity in a given cell. But when it is rotating in the opposite direction, the same cell is much less active. Because of the stereoscopic depth cues added to the dots, the cell is sensitive to the three-dimensional structure of the cylinder. In the bistable state, however, in which the cylinder is projected on a two-dimensional surface and there is no depth information, the monkey still tells us the direction he thinks the cylinder is rotating. Sometimes he says it is rotating one way, sometimes the other. When he thinks it is rotating clockwise, the nerve cell reliably reports this by the activity it generates corresponding to its perception. This result indicates that we have tapped into the area of the cortex that is analyzing this depth from motion, and we can actually see in the nerve-cell activity what the monkey is perceiving. And even though the information on his retina remains the same, the cells respond differently, indicating that the changes in perception – of which way the cylinder is turning – are occurring in this part of the brain (Figure 8).

If we continue upward along the motion pathway's hierarchical organization we come to a tiny area called Media Superior Temporal Area (MST), which is about half the size of the nail on your pinkie finger. Humans and monkeys both have an MST; it is specialized for helping us to navigate through the world using motion information. While you are driving along a highway or walking along a street, you generate motion signals. These signals are called optical flow. At the point – or focus – toward which you are headed,

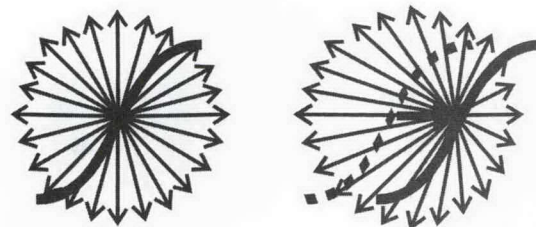


Fig. 10. MST neurons shift their focus tuning to compensate for eye movements

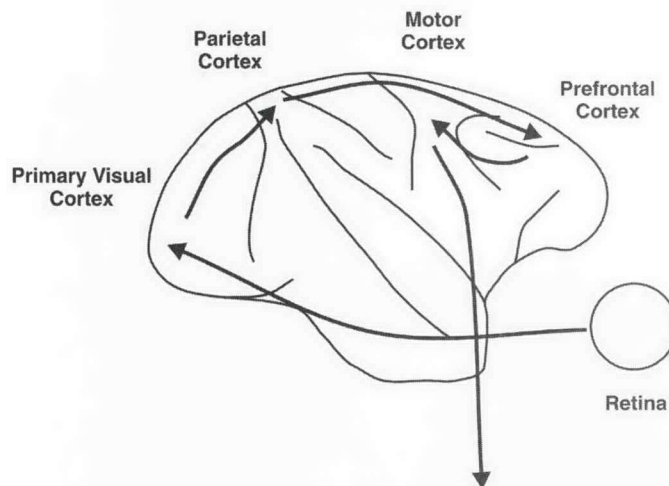


Fig. 11. Pathway of information that leads to visually-guided movements

there is very little motion, but around this focus point motion appears to radiate out, speeding out toward the edges of the visual field like an expanding circle. We call this spot the focus of expansion; it corresponds to the direction in which you are heading, and it gives you useful information about where you are going in the world. Cells in MST are tuned to these sorts of expanding stimuli generated by motion and also to the location of the focus. Now, a problem occurs when you are moving through the world in one direction but you begin to track something with your eyes – say a freeway sign – that may be off to the side. Moving your eyes introduces a motion of your visual field in the opposite direction. For example, if you hold a finger in front of you and follow it with your eyes as you move it to the right, you will notice that everything behind it moves to the left. With a rightward eye movement, you have introduced a leftward motion onto the eye. If you are also moving at the same time, this retinal motion gets combined with the expansion signal, shifting the focus toward the direction in which the eye is moving. If our brains were, in fact, using only this new focus to guide us through the world, when we looked at a sign on the freeway, we'd run into it, because that would be the point where the image is stabilized, with everything else radiating out from it.

But we know we don't do that. To find out what is going on in the brain during this process, we (David Bradley, Marsha Maxwell, and Krishna Shenoy from my lab; Marty Banks, a professor at UC Berkeley; and I) have recorded

from nerve cells in MST. The tuning curve of such a cell (which describes the frequency of the electrical signal coming from a cell) for an expanding stimulus is shown in Figure 10. If the expansion point is straight ahead, this cell is firing at about half activity; if the expansion point is over to the right, the cell is very active, and if it is to the left, the cell is not active at all. If we then have the monkey move its eyes so that it shifts the focus in the direction of the eye movement (the equivalent of looking at the freeway sign), we find that the nerve cells shift their tuning curves to compensate for the eye movement. The cell continues to fire at half activity, indicating that the monkey knows it has not changed its heading. What we think is happening is that the areas in the front part of the brain that are sending out signals to move the eyes are also sending signals back into the perceptual areas saying: *"The eye is moving; shift your receptive fields to compensate for it so that you still perceive locations in the world as being the same."* This mechanism is called efference copy or corollary discharge, and it explains why, when we move our eyes around and shift the images on our eyes, the world still appears stable. We are using information about what we are doing with our eyes to stabilize the visual world. Thus we can see that there is a hierarchy from V1, which measures motion, to MT, which extracts the 3-D structure of surfaces in motion, to MST, which helps us navigate through the world.

The final processing strategy that I will discuss is association. The bouncing red ball has now been divided up so that it is processed along three different streams – motion, color, and shape. But since we view the world as a unitary entity, at some point we need to begin bringing this information back together again into one picture. This binding of features back together occurs at the highest levels of the visual cortex, in the visual association areas.

A few years ago, our lab described an area called Lateral Interparietal Area (LIP), which is important for perceiving visual space and is located in the upper "where" processing stream. LIP is also important for making eye movements by gathering information from the visual cortex and sending it to the front part of the brain to move the eyes. However, we not only move our eyes to locate visual stimuli, but also to identify auditory stimuli. We know that our brains can perceive a sound location as easily as a visual location, but auditory information is collected in a very different way. It is assembled from auditory cues arriving at the two ears, while visual information is imaged on the retinas in the eyes. The brain has to combine these two very different types of signals to come up with a single, unified spatial representation. To this point, we tested LIP neurons only with visual signals. We were, however, interested in how this high-level processing area might combine or "associate" features of external stimuli to locate them in space. So we developed an auditory localization task.

It turned out that when Brigitte Stricanne, Pietro Mazzoni, and I recorded from nerve cells in the LIP area (which is a part of the posterior parietal cortex), we could also map

tuning curves or receptive fields for auditory stimuli. We had the monkey sit in a room with his head facing straight ahead, keeping his head always in the same position. He did, however, have to move his eyes to look at three different locations in the room. We played tones sequentially from speakers in different locations in order to map the cell's preferred location in space. When the animals looked in the three different directions, the preferred auditory location actually shifted in space by the same amount as the shift in gaze direction. In other words, the selectivity of the cell to the sound moves with the eye. This finding shows that the auditory signals have been mapped onto the same coordinate frame as the visual signals, which also move with the eyes. We say that both the auditory and visual signals are in an eye-centered reference frame. Auditory and visual information were brought together and associated in LIP to form a single common perceptual world representation.

In the last couple of years we have begun to investigate how sensory signals lead to decisions and plans for action. Working in such a high-order area as the posterior parietal cortex, with so many fascinating neural activities, we have wondered if intentions might be hatched here. Since the posterior parietal cortex lies between sensory areas and motor areas and acts as an interface between them, it seemed a likely candidate for the location of the neural correlates of intention. In experiments published in March in *Nature*, Larry Snyder, Aaron Batista, and I trained our monkeys to do one of two tasks when directed by a signal. On a green signal light they were to reach in the dark for the remembered location of a briefly flashed target; a red signal light told them to make an eye movement (saccade) to the target instead. They had to memorize the target's location over a delay of one to one and a half seconds before they acted. We measured the activity of specific neurons during this delay and discovered that the neurons fired not only to a specific location in the visual field but also according to whether the monkey was planning to look at or reach for the target. Moreover, the cells selective for eye movements were confined to area LIP, the saccade area, and the reach-selective cells were confined to a reach area abutting LIP. This anatomical segregation shows that a motor plan, guided by the visual perception, originates here in the culmination of the "where" pathway, and that the intended response, rather than the visual information, may be the determining factor in organizing how neural computations are made within the area. This may be the place where our thoughts begin to turn into actions, and where our spatial perception is mapped not only by what our senses tell us but also by how we plan to use that information.

ACKNOWLEDGEMENTS

I thank David Bradley for discussion and help with figures, and Sylvie Gertmenian for editorial assistance. //